Micromonospora echinospora genes ~~encoding for biosynthesis of~~ coding for biosynthesis of calicheamicin and self-resistance thereto

This application is a continuation-in-part of the non-provisional application 09/457045, filed December 7, 1999 and claims benefit thereof, which application is incorporated herein by reference in its entirety. This application also claims benefit from provisional application 60/111,325 filed on December 7, 1998, which application is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to a biosynthetic gene cluster of *Micromonospora echinospora* spp. *calichensis*. In particular, the calicheamicin biosynthetic gene cluster contains genes encoding for proteins and enzymes used in the biosynthetic pathway and construction of calicheamicin's aryltetrasaccharide and aglycone, and the gene conferring calicheamicin resistance. The present invention also relates to isolated genes of the biosynthetic cluster and their corresponding proteins. In addition, the invention relates to DNA hybridizing with the calicheamicin gene cluster and the isolated genes of that cluster. The invention also relates to expression vectors containing the biosynthetic gene cluster, the individual genes, or functional variants thereof.

Background of the Invention

The enediyne antibiotics, which were discovered in the 1980's, have long been appreciated for their novel molecular architecture, their remarkable biological activity, and their fascinating mode of action. Enediyne antibiotics were originally derived by

fermentation of microorganisms, including *Micromonospora, Actinomadura,* and *Streptomyces.* Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents,* p. 2 (1995). As a class, the enediyne antibiotics have been referred to as the most potent and highly active antitumor reagents yet discovered. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents,* preface (1995).

To date, at least twelve members of this family of antibiotics have been discovered, all of which fall roughly into two categories. The members of the first category of enediynes are classified as chromoprotein enediynes because they possess a novel 9-membered ring chromophore core structure, which also requires a specific associated protein for chromophore stabilization. The members of the second category of enediyne are classified as non-chromoprotein enediynes. These enediynes contain a 10-membered ring, which requires no additional stabilization factors. This enediyne ring structure is often referred to as the "warhead." The warhead induces DNA damage, which is frequently a double-stranded cleavage and appears to be irreparable. This type of DNA damage is usually nonrepairable for the cell and is most often lethal. Because of these remarkable chemical and biological properties, there has been an intense effort by both the pharmaceutical industry and academia to study these substances with the goal of developing new and clinically useful therapeutic anti-tumor agents.

The 9-membered ring chromoprotein enediyne subfamily is comprised of: neocarzinostatin from *Streptomyces carzinostaticus,*(Myers, A.G., et al., *J. Am. Chem. Soc., 110,* 7212-7214 (1988)); kedarcidin from *Actinomycete* L585-6, (Leet, J.E., et al., *J. Am. Chem. Soc., 114,* 7946-7948 (1992)), N1999A2 from *Streptomyces globisporus*, (Yoshida, K., et al. *Tetrahedron Lett., 34,* 2637-2640 (1993)), maduropeptin from

Actinomadura madurea, (Schroeder, D.R., et al., *J. Am. Chem. Soc.*, 116, 9351-9352 (1994)); N1999A2 from *Streptomyces sp.* AJ9493, (Schroeder, D.R., et al., *J. Am. Chem. Soc., 116,* 9351-9352 (1994)); actinoxanthin from *Actinomyces globisporus*, (Khokhlov, A.S., et al., *J. Antibiot., XXII,* 541-544 (1969)); largomycin from *Streptomyces pluricolorescens*, (Yamaguchi, T., et al., *J. Antibiot., XXIII,* 369-372 (1970)); auromomycin from *Streptomyces macromomyceticus,* (Yamashita, T., et al., *J. Antibiot., XXXII,* 330-339 (1979)), and sporamycin from *Streptosporangium pseudovulgare,* (Komiyama, K, et al., *J. Antibiot.,* XXX, 202-208 (1977)), all of which are believed to possess a novel bicylo[7.3.0.]dodecadiynene chromophore core structure essential for biological activity. In addition, with the exception of N1999A2, a required apoprotein acts as a stabilizer and specific carrier for the unstable chromophore, and for its transport and interaction with target DNA.

The non-chromophore enediyne subfamily is comprised of calicheamicin from *Micromonospora echinospora* spp. *calichensis*; namenamicin from *Polysyncraton lithostrotum;* esperamicin from *Actinomadura verrucosospora;* and dynemicin from *Micromonospora chersina.*

Enediyne antibiotics have potential as anticancer agents because of their ability to cleave DNA; however, many of these compounds are too toxic to be used currently in clinical studies. Today, only calicheamicin is known to be currently used in clinical trials; and it has provided promising results as an anticancer agent. For example, MyloTarg™, a calicheamicin-antibody conjugate also known as CMA-676 was approved by the FDA in January of 2000 to treat acute myelogenous leukemia. The enediynes also potentially have utility as anti-infective agents, provided that toxicity can be managed.

Calicheamicin has two distinct structural regions: the aryltetrasaccharide and the aglycone (also known as the warhead). The aryltetrasaccharide displays a highly unusual series of glycosidic, thioester, and hydroxylamine linkages and serves to deliver the drug primarily to specific tracts (5'-TCCT-3' and 5'-TTTT-3') within the minor groove of DNA when those sequences are available. However, specificity is also context-dependent. The aglycone of calicheamicin consists of a highly functionalized bicyclo[7.3.1]tridecadiynene core structure with an allylic trisulfide serving as the triggering mechanism. McGahren, W.J.,et al., *Enediyne Antibiotics as Antitumor Agents*, pp. 75-86 (1995). Once the aryltetrasaccharide is firmly docked, aromatization of the bicyclo[7.3.1]tridecadiynene core structure, via a 1,4-dehydrobenzene-diradical, results in the site specific oxidative double strand scission of the targeted DNA. Zein, N., et al., *Science*, 240, 1198-1201 (1988). The aglycone undergoes a reaction that yields carbon-centered diradicals, which are responsible for DNA cleavage.

This activity of calicheamicin has sparked considerable interest in the pharmaceutical industry culminating in the recent FDA approval of the calicheamicin-antibody conjugate MyloTarg™ (CMA-676) to treat acute myelogenous leukemia (AML). Additionally, similar strategies have been used in phase I trials to treat breast cancer. A massive program to examine calicheamicin conjugated to alternative delivery systems has also recently been undertaken. Hamann, P.R., et al., *87th Annual Meeting of the American Association of Cancer Research,* Washington, D.C., pp. 471 (1996); Hinman, L.M., et al., *Cancer Res., 53,* 3336 (1993); Hinman, L. M., et al., *Enediyne Antibiotics as Antitumor Agents,* pp. 87- 105 (1995); Sievers, E.L., et al., *Blood, 93,* 3678-3684 (1999); Siegel, M.M., et al., *Anal. Chem., 69,* 2716-2726 (1997); Ellestad, G. personal communication.

The biological activity and molecular architecture of calicheamicin has also prompted a search for potentially useful analogs. Of the numerous laboratories producing synthetic analogs, one group has produced a novel calicheamicin γ^{I}_{1} shown to effectively suppress growth and dissemination of liver metastases in a syngeneic model of murine neuroblastoma. Lode, H. N., et al., *Cancer Res., 58,* 2925-2928 (1998); Wrasidlo, W., et al., *Acta Oncologica, 34,* 157-164 (1995). In addition to synthesizing calicheamicin analogs, random mutagenesis of *M. echinospora* and screening for mutant strains with improved biosynthetic potential has also been pursued. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents,* pp. 107-126 (1995).

The first total synthesis of calicheamicin was reported by Nicolaou and coworkers in 1992. Synthesizing this complex antibiotic, though, presents many disadvantages. For example, Nacelle's procedure only provides approximately a 0.007% yield and requires 47 steps. Halcomb, R.L., Enediyne Antibiotics as Antitumor Agents, pp. 383-439 (1995). Thus, the total synthesis of calicheamicin remains secondary to the isolation of calicheamicin from large fermentations of *M. echinospora.* Therefore, methods to produce mass amounts of calicheamicin and potentially useful variants are still needed. Fantini, A., et al., *Enediyne Antibiotics as Antitumor Agents,* pp. 29-48 (1995). Transforming calicheamicin DNA into producing strains of bacteria, such as *Streptomyces*, *Micromonospora*, other actinomyces species, or *E. coli*, as non-limiting examples, would address this need. However, prior to the discoveries of the present inventors, no cloned *M. echinospora* genes were available, and only a set of limited studies upon putative *M. echinospora* promoters were available. Lin, L.S., et al., *J. Gen. Microbiol., 138,1881*-1885

(1992); Lin, L.S., et al., *J. Bacteriol.*, 174, 3111-3117 (1992); Baum, E.Z., et al., J. *Bacteriol.*, 171, 6503-6510 (1989); Baum. E.Z., et al., *J. Bacteriol.*, 170, 71-77 (1988).

Sub B2

Calicheamicin's molecular architecture in conjunction with its useful biological activity and potential therapeutic value brand calicheamicin an target for the study of natural product biosynthesis. While the radical-based mechanism of oxidative DNA cleavage by calicheamicin (i.e. aromatization of the bicyclo[7.3.1]tridecadiynene core structure, via a 1,4-dehydrobenzene-diradical, resulting in the site specific oxidative double strand DNA cleavage) is well understood, it was unknown, prior to this invention, how *Micromonospora* constructs calicheamicin. As a result, before the present invention, there was a need to discover and understand calicheamicin biosynthesis. Prior to this discovery of the present inventors, knowledge of genes encoding for nonchromoprotein enediyne biosynthesis was completely lacking.

Sub B3

The toxicity of the enediyne compounds, including calicheamicin, centers on the problem of directing the compound to the cleave only the DNA of interest, such as tumor cell DNA, and not the DNA of the host. Due to calicheamicin's powerful ability to cleave DNA, scientists have investigated the mechanism by which calicheamicin-producing organism protects itself against the DNA-cleaving activity of the molecule. Rothstein, D. M., *Enediyne Antibiotics as Antitumor Agents,* p. 77 (1995). Prior to this invention, knowledge of genes encoding for non-chromoprotein enediyne self resistance was completely lacking.

Summary of the Invention

The present invention relates to the first identification, isolation, and cloning of a nonchromoprotein enediyne biosynthetic gene cluster and mapping and nucleotide

sequence analysis of the genes within the cluster. The invention provides the entire calicheamicin-biosynthetic cluster and biochemical studies of aryltetrasaccharide biosynthesis. Furthermore, the calicheamicin self-resistance gene and protein have been isolated, as have the genes and resulting enzymes for steps within the calicheamicin cascade. The invention also provides for construction of enediyne overproducing strains, for rational biosynthetic modification of bioactive secondary metabolites, for new drug leads, and for an enediyne combinatorial biosynthesis program.

The present invention provides an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora* comprising said nucleic acid molecule, a portion or portions of said nucleic acid molecule wherein said portion or portions encode a protein, a portion or portions of said nucleic acid molecule wherein said portion or portions encode a biologically active fragment of a protein. The isolated nucleic acid molecule may be single- or double-stranded. As used herein, a nucleic acid molecule, polypeptide, or protein described as being "from" e.g., an organism or gene cluster, may have been isolated from such organism or gene cluster; alternatively, it may be a molecule which has been produced using synthetic, chemical, recombinant, or other such methods and comprise an amino acid or nucleotide sequence which may be isolated from such organism or gene cluster.

The present invention provides forty-eight genes, twenty-seven of which encode structural genes with the remainder encoding a variety of functions. The present invention is drawn to the following genes or nucleic acids: *calC* (SEQ ID No. 1), *calH* (SEQ ID No. 3), *calG* (SEQ ID No. 5), *calA* (SEQ ID No. 7), *calB* (SEQ ID No. 9), *calD* (SEQ ID No. 11, *calF* (SEQ ID No. 13), *calI* (SEQ ID No. 15), *calJ* (SEQ ID No. 17), *calK* (SEQ ID

No. 19), *calL* (SEQ ID No. 21), *calM* (SEQ ID No. 23), *calN* (SEQ ID No. 25), *calO* (SEQ ID No. 27), *calP* (SEQ ID No. 29), *calQ* (SEQ ID No. 31), *calR* (SEQ ID No. 33), *calS* (SEQ ID No. 35), *calT* (SEQ ID No. 37), *calU* (SEQ ID No. 39), *calV* (SEQ ID No. 41), *calW* (SEQ ID No. 43), *calX* (SEQ ID No. 45), *6MSAS* (SEQ ID No. 47), *ActI* (SEQ ID No. 49), *ActII* (SEQ ID No. 51), *ActIII* (SEQ ID No. 53), *orf1* (SEQ ID No. 55), *orf2* (SEQ ID No. 57), *orf3* (SEQ ID No. 59), *orf4* (SEQ ID No. 61), *orf5* (SEQ ID No. 63), *orf6* (SEQ ID No. 65), *orf7* (SEQ ID No. 67), *orf8* (SEQ ID No. 69), *orfI* (SEQ ID No. 71), *orfII* (SEQ ID No. 73), *orfIII* (SEQ ID No. 75), *orfIV* (SEQ ID No. 77), *orfV* (SEQ ID No. 79):, *orfVI* (SEQ ID No. 81), *orfVII* (SEQ ID No. 83), *orfVIII* (SEQ ID No. 85), *orfIX* (SEQ ID No. 87), *orfX* (SEQ ID No. 89), *orfXI* (SEQ ID No. 91), *IS-element* (DNA) (SEQ ID No. 93), *calE* (SEQ ID No. 94). The invention is also drawn to the following proteins or putative proteins: CalC (SEQ ID No. 2), CalH (SEQ ID No. 4), CalG (SEQ ID No. 6), CalA (SEQ ID No. 8), CalB (SEQ ID No. 10), CalD (SEQ ID No. 12), CalF (SEQ ID No. 14), CalI (SEQ ID No. 16), CalJ (SEQ ID No. 18), CalK (SEQ ID No. 20), CalL (SEQ ID No. 22), CalM (SEQ ID No. 24), CalN (SEQ ID No. 26), CalO (SEQ ID No. 28), CalP (SEQ ID No. 30), CalQ (SEQ ID No. 32), CalR (SEQ ID No. 34), CalS (SEQ ID No. 36), CalT (SEQ ID No. 38), CalU (SEQ ID No. 40), CalV (SEQ ID No. 42), CalW (SEQ ID No. 44), CalX (SEQ ID No. 46), 6MSAS (SEQ ID No. 48), ActI (SEQ ID No. 50), ActII (SEQ ID No. 52), ActIII (SEQ ID No. 54), Orf1 (SEQ ID No. 56), Orf2 (SEQ ID No. 58), Orf3 (SEQ ID No. 60):, Orf4 SEQ ID No. 62), Orf5 (SEQ ID No. 64), Orf6 (SEQ ID No. 66), Orf7 (SEQ ID No. 68), Orf8 (SEQ ID No. 70), OrfI (SEQ ID No. 72), OrfII (SEQ ID No. 74), OrfIII (SEQ ID No. 76), OrfIV (SEQ ID No. 78), OrfV (SEQ ID No. 80), OrfVI

(SEQ ID No. 82), OrfVII (SEQ ID No. 84), OrfVIII (SEQ ID No. 86), OrfIX (SEQ ID No. 88), OrfX (SEQ ID No. 90), OrfXI (SEQ ID No. 92), CalE (SEQ ID No, 95).

In one aspect, the present invention is directed to an isolated nucleotide molecule, wherein the nucleotide molecule hybridizes with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94, or a functional derivative of the isolated nucleotide molecule which hybridizes with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94. In one embodiment of the invention, the isolated nucleotide molecule has the nucleotide sequence of at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94, i.e., 100% complementarity (sequence identity) with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94. In another embodiment of the invention, the isolated nucleotide molecule has at least 90% complementarity (sequence identity) with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94. In yet another embodiment of the invention, the isolated nucleotide molecule has at least 80% complementarity (sequence identity) with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89,

91, 93 or 94. In yet another embodiment of the invention, the isolated nucleotide molecule has at least 70% complementarity (sequence identity) with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94. In yet another embodiment of the invention, the isolated nucleotide molecule has at least 60% complementarity (sequence identity) with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94. In still yet another embodiment of the invention, the isolated nucleotide molecule is substantially complementary to at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 or 94.

In another embodiment of the invention, there is provided an isolated protein encoded by a DNA molecule as described herein above, or a functional derivative thereof. A preferred protein has the amino acid sequence of at least one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, or 95 or a functional variant or derivative of one or more of those polypeptides.

In another embodiment, the present invention provides an isolated nucleic acid molecule from *Micromonospora echinospora* comprising a nonchromoprotein enediyne biosynthetic gene cluster, a portion or portions of said gene cluster wherein said portion or portions encode a protein, a portion or portions of said gene cluster wherein said portion or portions encode a biologically active fragment of a protein, a single-stranded nucleic

acid molecule derived from said gene cluster, or a single-stranded nucleic acid molecule derived from a portion or portions of said gene cluster.

In particular, the present invention provides an isolated nucleic acid molecule from *Micromonospora echinospora* spp. *calichensis* that is involved in the biosynthesis of calicheamicin. In another embodiment, the present invention also relates to nucleic acids capable of hybridizing with one or more isolated nucleic acids from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora* spp. *calichensis*. In a further embodiment, the invention provides an expression vector comprising an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora*. In yet a further embodiment the invention provides a cosmid comprising an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora*.

In preferred embodiments, the invention provides the isolated nucleic acid molecules of SEQ ID Nos. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 and 94.

In an additional embodiment, the present invention provides a host cell transformed with an isolated nucleic acid molecule from a nonchromoprotein enediyne biosynthetic gene cluster from *Micromonospora echinospora*. Host cells can optionally be of bacterial, yeast, fungal, insect, plant or mammalian origin and can be transformed according to standard methods. In a preferred embodiment, the host cell is the bacterium *E. coli, Streptomyces spp.*, or *Micromonospora spp.* In a more preferred embodiment, the host cell is the bacterium from the genus *Streptomyces* or from the genus *Micromonospora*.

In a further embodiment, the invention is directed to a host cell transformed with an expression vector comprising at least one of the nucleotide sequences of SEQ ID Nos. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 , or 94 or a portion of portions thereof or an allele or alleles thereof. In preferred embodiments, the host cells produce a biologically functional protein or portion of a protein, which protein or portion thereof is encoded by the expression vector.

In a specific embodiment, the invention is directed to a host cell transformed with an expression vector comprising *cal*C, or a portion(s) or allele(s) thereof, operably linked to regulatory sequences that enable expression of CalC. In another specific embodiment, the invention provides a host cell transformed with an expression vector comprising *cal*H, or a portion(s) or allele(s) thereof, operably linked to regulatory sequences that enable expression of CalH. In a yet further specific embodiment, the invention provides a host cell transformed with an expression vector comprising *cal*Q, or a portion(s) or allele(s) thereof, operably linked to regulatory sequences that enable expression of CalQ. Likewise, the invention provides a host cell transformed with an expression vector comprising *cal*G, or a portion(s) or allele(s) thereof, operably linked to regulatory sequences that enable expression of CalG.

In a yet further embodiment, the invention is directed to a host cell transformed with an expression vector encoding at least one polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, or 95 or a functional variant of one or more of those polypeptides. In

preferred embodiments, the host cells produce a biologically functional protein or portion of a protein, which protein or portion thereof is encoded by the expression vector.

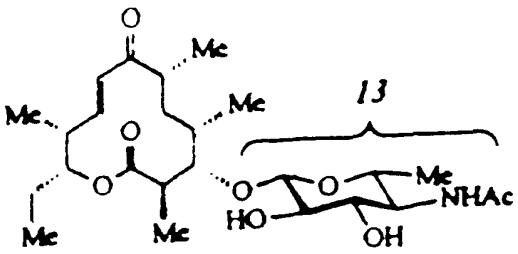
In a specific embodiment, the invention is directed to a host cell transformed with an expression vector encoding CalC, or a functional derivative thereof, operably linked to regulatory sequences that enable expression the encoded polypeptide. In another specific embodiment, the invention provides a host cell transformed with an expression vector encoding CalH, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of the encoded polypeptide. In a yet another specific embodiment, the invention provides a host cell transformed with an expression vector encoding CalQ, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of the encoded polypeptide. Likewise, the invention provides a host cell transformed with an expression vector encoding the CalG, or a functional derivative thereof, operably linked to regulatory sequences that enable expression of the encoded polypeptide.

The invention further provides a method of expressing a protein by culturing a host cell transformed with an expression vector of the present invention, and incubating the host cell for a time and under conditions allowing for protein expression.

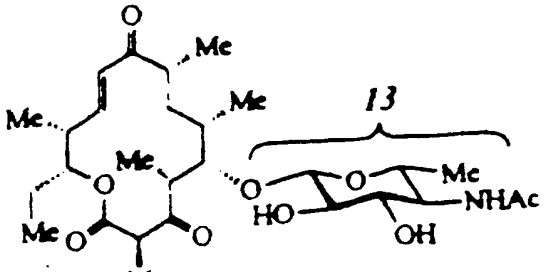
In yet another embodiment the invention provides a method of purifying calicheamicin using affinity chromatography. A sample containing calicheamicin is contacted with an affinity matrix having the protein CalC bound thereto, for a time and under conditions allowing calicheamicin to bind to the matrix, eluting calicheamicin from the matrix, and recovering calicheamicin.

In a further embodiment the present invention provides polypeptides comprising the amino acid sequences of SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95.

In yet a further embodiment the invention provides the production of the following two new macrolides:



11



12

The invention further provides a method of conferring calicheamicin resistance to a subject comprising obtaining cells from the subject, transforming the cells with the calicheamicin self-resistance gene, and returning the cells to the subject. Alternatively, the calicheamicin self-resistance gene can be targeted and delivered to the desired host cells through known gene therapy delivery systems.

The invention further provides a method of producing calicheamicin analogs by altering calicheamicin or its bioactive metabolites through the modulation of the expression of *calD, E, F, G, H, J, K, N, O, P, Q, S, T, U, V, W, X, 6MSAS, actI-III, orfI, orfIII, orfV*, and *orfVII.* Such modulation can be achieved through selective "knock out",

as well as heterologous expression of these genes and their products. Various combinations of these either mutated or wild type gene products may be used in either *in vitro* or *in vivo* calicheamicin analog production.

The invention further provides a method for increasing the production of calicheamicin through the introduction of multiple copies of positive regulators and transporters and or by eliminating or reducing the expression of negative regulators (e.g., CalA, B, I, L, Orf8). Additionally, upregulation of calicheamicin resistance genes *calC, calN* and *orfXI* can be used to decrease the toxicity of calicheamicin to healthy tissues and cells during therapy.

In a yet further embodiment, the invention provides for a method of transposon mediated mutagenesis or moving chromosomal DNA fragments *in vivo* through expression of the *orf3* integrase and the IS insertional element.

The advantages of the present invention are numerous. Isolation of and the ability to clone calicheamicin DNA opens the door for genetic analysis of calicheamicin biosynthesis, as such analysis requires the ability to obtain large quantities of DNA which codes for calicheamicin biosynthesis. Using the teachings of the present invention, one can study calicheamicin biosynthesis via mutagenesis of *M. echinospora*. For example, one can isolate and characterize mutants blocked in calicheamicin biosynthesis and then analyze their defective or partial calicheamicin products. Additionally, particular a enzyme or enzymes can be overexpressed or underexpressed after subcloning its gene into a host such as *E. coli*, and the results of such overexpression or underexpression can be studied to reveal the enzyme's function. Furthermore, the cloning of biosynthetic genes

can ultimately result in increased yields of the gene product by cloning and expressing the biosynthetic gene encoding the rate-limiting enzyme back into the producing organism.

Further, it may also be possible to generate novel products by cloning biosynthetic genes into strains that make related compounds. Such genes could endow the host organism with the ability to carry out new reactions on the enediyne nucleus, and thus produce novel drugs. The present invention thus also provides means for biosynthetic modification of bioactive secondary metabolites through enediyne combinatorial biosynthesis. As most pharmaceutical drug leads are inspired by naturally occurring compounds, and given the challenge posed in synthesizing these metabolites, genetic manipulation of the sugar appendage on the metabolites offers avenues for creating potential new drugs. Thus the emerging field of combinatorial biosynthesis has become a rich new source for modified non-natural sugar scaffolds. Marsden, A., et al., *Science* 1998, 279, 199-201. Problems inherent with the genetic manipulation of the sugar appendage relate to the fact that naturally occurring bioactive secondary metabolites possess unusual carbohydrate ligands, which serve as molecular recognition elements critical for biological activity. *Macrolide Antibiotics, Chemistry, Biology and Practice, 1984.* Without these essential sugar attachments, the biological activities of most clinically important secondary metabolites are either completely abolished or dramatically decreased. Currently, techniques for the genetic manipulation of the sugar appendage for a given metabolite rely mainly on the alteration and/or deletion of a small subset of genes required to construct and attach each desired sugar moiety. Thus there is a need to develop alternate strategies to construct and attach non-naturally occurring sugars. The present invention addresses this need. The present invention utilizes the fact that

glycosyltransferases, which are responsible for the final glycosylation of certain secondary metabolites, show a high degree of promiscuity toward the nucleotide sugar donor. Zhao, L., et al., *J. Am. Chem. Soc.* 1988, 120, 12159-12160. This unselectivity of the glycosyltransferases has the potential for allowing modification of the crucial glycosylation pattern of natural, or non-natural, secondary metabolite scaffolds in a combinatorial fashion. The present invention discloses a method using the recruitment and collaborative action of sugar genes from a variety of biosynthetic pathways to construct composite gene clusters, which make and attach non-natural sugars.

Insight into how *Micromonospora* self resistance gene and gene products act to control the toxic effects of calicheamicin offers new avenues of clinical research. For example, knowledge of the mechanisms underlying calicheamicin resistance, as provided by the present disclosure, can provide the means necessary to use higher doses of calicheamicin while simultaneously inhibiting the toxic effects of the drug on non-cancer cells. Additionally, understanding the mechanism behind calicheamicin's self-resistance may aid in the understanding of self-resistance in other enediyne antibiotics, thereby potentially making useful those enediynes once thought to be too toxic to be viably used as therapeutic agents. The calicheamicin self-resistance mechanisms elucidated utilizing the present invention provide gene therapy approaches, for example, via introduction of enediynes resistance genes into bone marrow cells, thereby increasing resistance and allowing tolerance to chemotherapeutic doses of calicheamicin. Banerjee, D., et al., *Stem Cells,* 12, 378-385 (1994). Thus, understanding calicheamicin self-resistance will significantly aid continuing clinical studies involving calicheamicin and the enediynes. The present invention addresses this need as it provides for the isolation and

characterization of a resistance gene and its associated protein for any nonchromoprotein enediynes.

Brief Description of the Figures

Figure 1 depicts the summary of the cosmid clones isolated from *M. echinospora* genomic library. This figure illustrates the results of the screening of the genomic library for clones carrying the calicheamicin biosynthetic cluster.

Figure 2 shows a restriction map of a portion of cosmid clones 4b, 13a, and 56 and the corresponding location of *cal* genes from *M. echinospora.*

Figure 3 is a table of the open reading frames ("orfs") in the calicheamicin biosynthetic cluster. This table lists the polypeptides that the genes encode for as well as their proposed or actual determined function in the biosynthetic pathway. [a] Assignments based upon BLAST search at the amino acid level unless otherwise noted. [b] Highest probability score obtained. [c]Assignment based on biochemical studies. [d] Only a portion of the orf has been elucidated.

Figure 4 is a graph of the UV-visible absorption spectra of purified mbp-CalC. The purified mpb-CalC was analyzed in the following solution: 52 μM mpb-CalC; 10 mM Tris-HCl, pH 7.5). The inset shows the results of low temperature (4.3 K) the X-band EPR analysis of CalC. 250 μM mpb-CalC containing 0.5 mol Fe per mol CalC was analyzed in 10 mM Tris-HCl, pH 7.5. The spectrometer settings were as follows: field set = 2050 G; scan range = 4,000G; time constant = 82 s; modulation amplitude =16 G; microwave power = 31 μW; frequency = 9.71 Ghz; gain = 1000; determined spin quantitation = 90 ± 10 μM Fe.

Figure 4(b) provides the results of the mbp-CalC *in vitro* assay.

Figure 5 depicts the postulated routes for the biosynthesis of required nucleotide sugars. The enzymes are depicted as follows: E_{deox} = deoxygenase; E_{am} = aminotransferase; E_{ep} = epimerase; E_{met} = methyltransferase; E_{od} = 4,6-dehydratase; E_{ox} = oxidase; E_p = nucleotidyltransferase; E_{red} = reductase; E_{sh} = sulfhydrytransferase.

Figure 6 illustrates a schematic representation of the *in vivo* production of pikromycin/methymycin-calicheamicin hybrid metabolites.

Figure 7 depicts the *Streptomyces venezuela* methymycin/pikromycin gene cluster. Eight open reading frames (*des*I-*des*VIII) in this cluster have been assigned as genes involved in desosamine biosynthesis. This figure also depicts the hybrid pathway toward new methymycin/pikromycin derivatives (11 and 12) produced after heterologous expression of the *cal*H gene of calicheamicin in a *S. venezuela* mutant.

Figure 8 illustrates calicheamicin's (6) four unique sugars which are crucial to tight DNA binding. Sugar (9) is derived from 4-amino-4,6-dideoxyglucose (8) and is part of the restricted N-O connection between sugars A and B. Compound 8 is derived from the corresponding 4-ketosugar (7) via a transamination reaction. The gene *cal*H encodes the desired C-4 aminotransferase allowing conversion of compound (7) to compound (8).

Figure 9 is a map illustrating the relative loci of the 48 identified genes spanning approximately 65KB of continuous sequence. Eight of the genes identified show no homologs in the public databases.

Figure 10 depicts additional postulated routes for the biosynthesis of required nucleotide sugars. The enzymes are depicted as follows: E_{deox} = deoxygenase; E_{am} = aminotransferase; E_{ep} = epimerase; E_{met} = methyltransferase; E_{od} = 4,6-dehydratase; E_{ox} = oxidase; E_p = nucleotidyltransferase; E_{red} = reductase; E_{sh} = sulfhydrytransferase.

Figure 11 is a schematic showing the iodination of orsellenic acid mediated by CalV and CalT, as well as the subsequent steps of oxidation, mediated by CalS and CalW and methylation, mediated by CalD and CalJ. Additionally, the figure shows the synthesis of putative substrates for the reaction.

Figure 12 describes the mechanism of calicheamicin resistance in *Micromonospora*. *cal*C confers calicheamicin resistance to bacteria.

Figure 13 A schematic diagram of the first continuous assay for enediyne-induced DNA cleavage, the Molecular Break Lights. The solid lines represent covalent bonds, dashed lines represent hydrogen bonding, letters represent arbitrary bases, the gray shaded ball represents the fluorophore (FAM: fluorescein), the black ball represents the corresponding quencher (DABCYL:4-(4-'demethylaminophenylazo)-benzoic acid) and the dashed wedges represent fluorescence. Generally, molecular beacons operate by a separation of the fluorophore-quencher pair resulting in a corresponding fluorescent signal. Molecular break lights, as illustrated in the figure, operate through cleavage of the stem by an enzymatic or non-enzymatic nuclease activity resulting in the separation of the fluorophore-quencher pair and corresponding fluorescent signal. In this study, Molecular break lights contain either a preferred calicheamicin recognition site (bold-faced, TCCT) or the *BamHI* recognition site (bold-faced, GGATCC). The predicted cleavage sites are illustrated by arrows.

Figure 14 shows the demonstration of molecular break light specificity and general proof of principle. The observed change in fluorescence intensity over time of an assay containing 3.2 nM break light at 37 °C. **(a)** Break light calicheamicin MLB (break light A) with 100 U *BamHI* (❑),*BamHI* MLB (break light B) with 100 U *BamHI* (o) and

BamHI MLB without enzyme (•) (10 mM Tris·HCl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9; λ_{Ex} = 485 nm, λ_{Em} = 517 nM). **(b)** calicheamicin MLB (break light A) with and 10 U DNaseI (❑),*BamHI* MLB (break light B) with 10 U DNaseI (o) and calicheamicin MLB (break light A) without enzyme (•) (40 mM Tris·HCl, 10 mM $MgSO_4$, 1 mM $CaCl_2$, pH 8.0; λ_{Ex} = 485 nm, λ_{Em} = 517 nM). This is the most sensitive assay for *BamHI* and DNaseI DNA cleavage activity to date.

Figure 15 shows the cleavage of calicheamicin MLB (break light A) by calicheamicin and esperamicin. The observed DNA cleavage over time of an assay containing 3.2 calicheamicin MLB at 37 °C (40 mM Tris·HCl, pH 7.5; λ_{Ex} = 485 nm, λ_{Em} = 517 nM), DTT (50 μM) and varied enediyne. **(a)** Calicheamicin concentrations: 31.7 nM (o), 15.9 nM (❑), 3.2 nM (◇), 1.6 nM (Δ), 0.78 nM (•) and 0.31 nM (■). **(b)** Esperamicin concentrations: 31.7 nM (o), 15.9 nM (❑), 3.2 nM (◇), 1.6 nM (Δ), 0.78 nM (•), 0.31 nM (■) and 0.15 nM (♦). These results represent the first continuous and most sensitive assay for enediyne-induced DNA cleavage.

Figure 16 **(a)** The observed DNA cleavage over time of an assay containing a constant 3.2 nM break light A at 37 °C (50 mM sodium phosphate, 2.5 mM ascorbate, pH 7.5; λ_{Ex} = 485 nm, λ_{Em} = 517 nM) and varied bleomycin. Bleomycin concentrations: 200 nM (o), 100 nM ( ), 50 nM (◇), 25 nM (Δ), 12.5 nM (•), 5 nM (■) and 2.5 nM (▲).**(c)** The observed DNA cleavage over time of an assay containing a constant 32 nM break light A at 37 °C (40 mM Tris·HCl, 2.5 mM ascorbate, pH 7.5; λ_{Ex} = 485 nm, λ_{Em} = 517 nM) and varied MPE. Fe(II) concentrations: 50 nM (o), 125 nM (❑), 250 nM (◇), 500 nM (Δ), 1 μM (•) and 2 μM (■). **(d)** The observed DNA cleavage over

time of an assay containing a constant 32 nM break light A at 37 °C (40 mM Tris·HCl, 2.5 mM ascorbate, pH 7.5; λ_{Ex} = 485 nm, λ_{Em} = 517 nM) and varied Fe^{+2}-EDTA. Fe(II) concentrations: 12.5 μM (○), 6.3 M (❑), 3.1 μM (◇), and 1.3 μM (Δ).

Figure 17 shows the direct in vitro inhibition of calicheamicin-mediated DNA cleavage using the break light assay. 3.6pM break light A is coincubated with 3.5nM calicheamicin with increasing amounts of CalC. Complete inhibition of calicheamicin is achieved with roughly 2-fold excess of CalC. CalC has no effect on esperamicin-induced cleavage of DNA.

Figure 18 shows the interaction between CalC and "activated" calicheamicin as measured by an increase in tryptophan fluorescence of CalC. CalC has 5 tryptophan and no cysteine residues and is unaffected by the reductive activator dithiothreitol (DTT). As the concentration of calicheamicin (3) increases in the absence of DTT there is little change in the CalC Trp fluorescence intensity. The addition of DTT to "activate" calicheamicin (4) results in increased binding to CalC as shown by the increase in CalC Trp fluorescence intensity.

Detailed Description of the Invention

The present invention is directed to the isolation and characterization of the calicheamicin biosynthetic cluster. This cluster encodes the genes that encode the proteins and enzymes that are involved in deoxysugar synthesis (the aryltetrasaccharide), polyketide biosynthesis (the aglycone and aromatic residue of the aryltetrasaccharide) of calicheamicin synthesis, regulation, transport, cluster mobility and calicheamicin resistance. Forty-eight putative genes have been identified, twenty-seven of which encode

putative structural proteins with the remainder encoding a variety of functions. Specifically, there are 15 genes that encode for the aryltetrasaccharide moiety (20,928 bp; *D, E, F, G, H, J, K, N, O, Q, S, T, U, X, W, 6MSAS*), 12 putative genes which encode for the aglycone (13,284 bp; *P, S, V, W, ActI, ActII, ActIII, OrfI, OrfIII, OrfV, OrfVI, OrfVII*), 13 putative genes involved in membrane transport, regulation, DNA movement and/or resistance (19,704 bp; *A, B, C, I, L, M, R, orf4, orf8, OrfVIII, OrfIX, OrfX, OrfXI, IS-element*), and the remaining 8 genes of unknown function (7383 bp; *orf1, orf2, orf3, orf5, orf6, orf7, OrfII, OrfIV*).

The calicheamicin biosynthetic gene cluster comprises the following genes: *cal*A, *cal*B, *cal*C, *cal*D, *cal*E, *cal*F, *cal*G, *cal*H, *cal*I, *cal*J, *cal*K, *cal*L, *cal*M, *cal*N, *cal*O, *cal*P, *cal*Q, *cal*R, *cal*S, *cal*T, *cal*U, *cal*V, *cal*W, *cal*X, *6MSAS, ActI, ActII, ActIII, orf*1, *orf2*, *orf*3, *orf*4, *orf*5, *orf*6, *orf7*, *orf8*, *orfI*, *orfII orfIII*, *orfIV orfV*, *orfVI*, *orfVII, orfVIII, orfIX*, *orfX*, *orfXI* and an IS-element gene. It should be noted that orf1-8 may contain DNA derived in whole or in part from recombinant vectors LP46 and/or LP54. The above listed genes encode the following polypeptides: CalA (328 amino acids), CalB (561 amino acids), CalC (181 amino acids), CalD (263 amino acids), CalE (420 amino acids), CalF (245 amino acids), CalG (990 amino acids), CalH (338 amino acids), CalI (568 amino acids), CalJ (332 amino acids), CalK (440 amino acids), Cal L (562 amino acids), Cal M (416 amino acids), CalN (398 amino acids), CalO (331 amino acids), Cal P (approximately 179 amino acids), CalQ (453 amino acids), CalR (265 amino acids), CalS (1113 amino acids), CalT (280 amino acids), CalU (377 amino acids), CalV (125 amino acids), CalW (449 amino acids), CalX (197 amino acids), 6MSAS (198 amino acids), ActI (207 amino acids), ActII (136 amino acids), ActIII (308 amino acids), Orf1(322 amino acids), Orf2

(654 amino acids), Orf3 (209 amino acids), Orf4 (521 amino acids), Orf5 (175 amino acids), Orf6 (139 amino acids), Orf7 (187 amino acids), Orf8 (266 amino acids), OrfI (127 amino acids), OrfII (248 amino acids) OrfIII (298 amino acids), OrfIV (363 amino acids) OrfV (288 amino acids), OrfVI (1012 amino acids), OrfVII (236 amino acids), OrfVIII (441 amino acids), OrfIX (504 amino acids), OrfX (504 amino acids), OrfXI (251 amino acids) and IS-element (402 amino acids).

In elucidating the calicheamicin biosynthetic gene cluster, the inventors began with a genomic library containing the genome of *Micromonospora echinospora* spp. *calichensis*. The cosmid library was generated by isolating chromosomal DNA of *Micromonospora echinospora* spp. *calichensis*, fragmenting that chromosomal DNA, inserting the DNA into a cosmid vector and generating a cosmid library according to methods well known in the art. This procedure can be performed using any species of *Micromonospora, Streptomyces*, or other suitable bacteria.

Based upon prior enediyne metabolic labeling studies it was postulated that the calicheamicin aglycone would be polyketide derived. Polyketide metabolites encompass a vast variety of structural diversities yet share a common mechanism of biosynthesis. Hutchinson, C.R., et al., *Chem. Rev., 97,* 2525-2535 (1997); Strohl, W.R., et al, *Biotechnology of Antibiotics* pp. 577-657; Fujii, I., et al., *Chem. Rev., 97,* 2511-2523 (1997); Hopwood, D.A., et al., *Chem. Rev., 97,* 2465-2497 (1997); Hopwood, D.A., et al., *Ann. Rev. Genet., 24,* 37-66 (1990); Staunton, J., et al., *Chemical Reviews, 97,* 2611-2629 (1997). Most important, polyketide synthase ("PKS") genes display a high degree of sequence homology (from pathway to pathway and organism to organism) and are often clustered with genes encoding self resistance and deoxysugar ligand biosynthesis.

Hopwood, D.A., et al., *Chem. Rev., 97,* 2465-2497 (1997); Hopwood, D.A., et al., *Ann. Rev. Genet., 24,* 37-66 (1990); Staunton, J., et al., *Chem. Rev., 97,* 2611-2629 (1997).

Degenerate primers based upon conserved regions within PKS genes were used in Southern hybridizations to identify clones from the *M. echinospora* genomic library that carried putative PKS genes. The Southern hybridizations were performed by methods known in the art. Southern hybridization of the genomic *M. echinospora* cosmid library with a DNA probe designed to target type I PKS genes (KS^{I}), (Kakavas, S.J., et al., *J. Bacteriol., 179,* 7515-7522 (1997)), unveiled five positive clones, which were designated clones 4b, 10a, 13a, 56, and 60. See Figure 1. The same five clones were also identified upon rescreening the genomic library with type II DNA probe (actI). See Figure 1. Although this preliminary analysis clearly demonstrated the presence of *Micromonospora* PKS gene homologues, a secondary screen was performed, as PKS hybridization analyses are often plagued by false hybridization to gene clusters that encode spore pigment biosynthesis.

The second screening was based on the assumption that calicheamicin's biosynthetic cluster would also contain genes encoding for deoxysugar ligand synthesis. Further, it was postulated that all hexopyranosyl ligands of calicheamicin diverged from the common intermediate 4-keto-6-deoxy TDP-D-glucose (30), Figure 5, as macromolecule-sugar synthesis in many organisms began with a similar common intermediate. Thus, it was believed that the cluster encoding for calicheamicin biosynthesis, in addition to carrying a PKS-encoding region, would carry both a common glucose-1-phosphate nucleotidyltransferase and a NDP-α-D-glucose 4,6-dehydratase gene, encoding the putative enzymes E_{p1}, and E_{od}, respectively. See figure 5. These enzymes are

necessary to convert a sugar (12)(figure 5) to the hypothesized common intermediate, 4-keto-6-deoxy TDP-D-glucose (30). Analogs to 4,6-dehydratases have been previously characterized from *E. coli, Salmonella,* and *Streptomyces.* Additionally, a nucleotide transferase from *Salmonella* has been characterized as an alpha-D-glucose-1-phosphate thymidylyltransferase. The secondary screen was performed using a probe based upon the postulation that the *M. echinospora*'s calicheamicin synthesis would begin from a similar precursor found in *E. coli, Streptomyces* and *Salmonella*, and that this precursor required a dehydratase to convert it into the common intermediate, 4-keto-6-deoxy TDP-D-glucose (30). In particular, a DNA probe (designated E_{od}^{I}) was designed from the conserved NAD^{+}-binding site of bacterial NDP-α-D-glucose 4,6-dehydratases. He, X., et al., *Biochem., 35,* 4721-4731 (1996). Southern hybridization of the genomic *M. echinospora* cosmid library with the E_{od}^{I} probe revealed cross-hybridization with clones 4b, 10a, 13a, 56, and 60. Two additional clones, designated 58 and 66, were also identified in this screen. See Figure 1. This secondary hybridization indicated the clustering of genes encoding both polyketide and deoxysugar biosynthesis.

For final corroboration, since secondary metabolite biosynthesis is typically clustered with resistance genes in actinomycetes, all hybridization-positive clones were tested for their ability to grow in the presence of varying concentrations of calicheamicin. In this final screen, six of the seven hybridizing clones displayed differing levels of resistance to calicheamicin (4b≈10a≈13a≥56≥66>60)(See Figure 1) while clone 58 lacked the ability to grow in the presence of calicheamicin. In addition, these resistance screens revealed that clones 4b, 10a, 13a conferred much higher levels of resistance to calicheamicin than the other clones. Upon rescreening the genomic library for

calicheamicin-resistant clones, three additional clones (3a, 4a, and 16a) were found to confer similar levels of resistance. Cumulatively, the results demonstrated that clones 4b, 10a, 13a, 56, and 60 carried PKS I and II homologues and deoxy sugar biosynthetic genes, as well as encoded the gene responsible for conferring calicheamicin-self resistance.

The clones positive for PKS I and II and deoxy sugar biosynthesis homology and calicheamicin resistance were used to map the biosynthetic cluster. Southern hybridization established similarity between clones 3a, 4a, 4b, 10a, 13a, 16a and 56. In addition, nucleotide sequence overlaps were found between clones 4b, 13a, and 56. See Figure 1. Restriction mapping and Southern hybridization of these clones indicated that the positive cosmid clones corresponded to a continuous region of the *M. echinospora* chromosome spanning > 100 kb. The present invention thus provides for cosmids having a nucleic acid molecule from *Micromonospora echinospora* encoding for a nonchromoprotein enediyne biosynthetic cluster.

After isolating the biosynthetic gene cluster and elucidating the sequence, open reading frames ("orfs") were assigned. Tentative gene assignments were derived from amino acid sequence similarity of translated orfs to gene products of known function via direct BLAST (Basic Local Alignment Search Tool) database searches on the amino acid level. Karlin, et al., *Proceed Natl. Acad. Sci., U.S.A., 87,* 2264-2268 (1990); Karlin, et al., *Proceed Natl. Acad. Sci., U.S.A., 90,* 5873-5877 (1993); Altchul, *Nature Genet., 6,* 119-129 (1994). The gene cluster organization is provided in figure 1.

Based on BLAST analysis tentative gene assignments were made. Specifically, there are 15 genes that encode for the aryltetrasaccharide moiety (20,928 bp; *D, E, F, G, H, J, K, N, O, Q, S, T, U, X, W, 6MSAS),* 12 putative genes which encode for the aglycone

(13,284 bp; *P, S, V, W, ActI, ActII, ActIII, OrfI, OrfIII, OrfV, OrfVI, OrfVII),* 13 putative genes involved in membrane transport, regulation, DNA movement and/or resistance (19,704 bp; *A, B, C, I, L, M, R, orf4, orf8, OrfVIII, OrfIX, OrfX, OrfXI, IS-element*), and the remaining 8 genes of unknown function (7383 bp; *orf1, orf2, orf3, orf5, orf6, orf7, OrfII, OrfIV*).

One aspect of the invention relates to transformation of a host cell with *M. echinospora* DNA. This method provides a reproducible transformation efficiency of $\sim 10^3$ kanamycin resistant transformants/ g DNA using a pKC1139-based vector. The invention further provides that the host cell can be but is not limited to bacteria, yeast, fungus, insect, plant or mammalian. Transformations of bacteria, yeast, fungus, insect, plant or mammalian cells are performed by methods known in the art.

The present invention also provides the isolation and characterization of genes encoding polypeptides involved in calicheamicin resistance such as *orfXI* and *calC* . One aspect of the invention relates to an isolated DNA strand having the gene *cal*C and having the DNA sequence SEQ. ID No.: 1. The present invention also relates to an isolated protein CalC, having the amino acid sequence, SEQ ID. NO. 2. The invention further provides for *cal*C gene fragments coding for a bioactive CalC polypeptide. The polypeptide, CalC, confers calicheamicin resistance and has 181 amino acids. The invention also provides for CalC fragments conferring calicheamicin resistance.

The *cal*C locus was isolated by identifying calicheamicin genomic cosmid clones that were able to grow on luria bertani ("LB") agar plates containing ampicillin and calicheamicin. The DNA of the positive clones (clones that grew on the plates containing calicheamicin) was isolated and subsequent restriction mapping localized the desired

phenotype (calicheamicin resistance). The DNA was then sequenced and the open reading frames analyzed to ascertain the orf encoding for the desired phenotype. *In vitro* studies were also performed and confirmed the ability of CalC to inhibit DNA cleavage.

DNA containing *cal*C was cloned into an inducible vector, using known methods, resulting in overexpression of *cal*C. The polypeptide product (CalC) was then isolated and purified to homogeneity. Analysis of the purified CalC revealed that it is a non-heme iron metalloprotein that functions via inhibition of calicheamicin-induced DNA cleavage *in vitro*. Another aspect of the invention is an expression vector containing *cal*C or a fragment of *cal*C encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably *E. coli*, containing *cal*C or a fragment of *cal*C encoding for a bioactive molecule. Such transgenic expression of *calC* results in an 10^5-fold increase in calicheamicin resistance in *E.coli*, a 100-fold increase in resistance in *S.lividans*, and a 50-fold increase in resistance in yeast.

The present invention provides for the transformation of human cells with the *cal*C gene. The transgenic expression of *calC* in the HT1080 (human) cell line increased its resistance to calicheamicin 10-fold. This technique allows bone marrow cells, for example, to be removed from a patient being treated with calicheamicin, and for these cells to be transformed with *cal*C, and for the transformed cells to be returned to the patient. This allows the patient to tolerate treatment with calicheamicin or allows the patient to receive higher doses of calicheamicin as the returned human-*cal*C-transformed cells have calicheamicin resistance. The transformation is performed by methods known in the art. The embodiment of the invention would be applicable to many diseases being treated with calicheamicin.

The invention further provides for a method of assaying the calicheamicin-induced DNA cleavage and its CalC-mediated inhibition using the molecular break light assay. Two molecular break lights (MLBs) for the experiments are described in example 7. Break light A is comprised of a 10-base pair stem which contained the known calicheamicin recognition sequence 5'-TCCT-3', while break light B carries the *BamH*I endonuclease recognition sequence 5'-GGATCC-3'. The 5'-fluorophore of both probes was fluorescein (FAM, absorbance$_{max}$ = 485 nm, emission$_{max}$ = 517 nm) while the corresponding 3'-quencher was 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL). Generally, MLBs operate by a separation of the fluorophore-quencher pair resulting a corresponding fluorescent signal. The molecular break lights, as illustrated in figure 13, operate through cleavage of the stem by specific enzymatic or non-enzymatic nuclease activity resulting in the separation of the fluorophore-quencher pair and corresponding fluorescent signal (see figure 14). CalC in a two-fold molar excess of calicheamicin, completely abolishes calicheamicin mediated DNA cleavage as monitored by the break light assay (see figure 15).

CalC acts as a "cleavage sink". In essence the protein is cleaved as an alternative to the desired DNA target. Thus, the invention provides the first such demonstrated mechanism for resistance to a cleavage agent and explains why CalC is able to function in all organisms tested so far (i.e. *E.coli*, *S.lividans*, yeast, and humans).

The invention further provides for the use of the break light assay to determine calicheamicin titers during production of thereof. Furthermore, the molecular break light assay may be used to determine the DNA cleavage activity of calicheamicin analogs generated using the techniques of this invention.

Another aspect of the invention relates to an isolated DNA strand containing the *cal*H gene having the DNA sequence SEQ ID. No: 3. The invention also relates to the polypeptide CalH, having amino acid sequence SEQ ID. No. 4. The invention further provides for *cal*H gene fragments coding for a bioactive CalH. CalH is involved in the formation of the aryltetrasaccharide 4,6-dideoxy-4-hydroxylamino-D-glucose moiety. CalH catalyzes the conversion of intermediate (30) to intermediate (39) (figure 5). CalH is a TDP-6-deoxy-D-glycerol-L-threo-4-hexulose 4-transaminase, which catalyzes a pyridoxal phosphate ("PLP")-dependent transamination from glutamate to provide 4-amino-6-deoxy TDP-D glucose (intermediate 39)(figure 5). The invention also provides for CalH fragments that retain bioactivity. There is also provided an expression vector containing the *cal*H gene or fragments of the *cal*H gene that encode for a bioactive polypeptide. CalH were overexpressed as a $(histidine)_{10}$-fusion protein and subsequently purified by nickel affinity chromatography.

According to BLAST analysis, CalH closely resembles perosamine synthase, an enzyme which converts compound 30 to compound 39 (See figure 5) *en route* to the biosynthesis of TDP-perosamine (TDP-4,6-dideoxy-4-amino-D-mannose) in *E. coli.* Wang, L., et al., *Infect. Immunol., 66,* 3545-3551 (1998). Thus CalH is believed to be a 4-ketohexose aminotransferase. To confirm the tentative BLAST assigned function, a combinatorial biosynthesis was performed. Specifically the *cal*H gene from calicheamicin was incorporated into a mutant strain of *Streptomyces venezuela.* The 4-dehydrase gene (*des*1) in the methymycin/pikromycin pathway was deleted in this mutant strain. A promoter sequence from the *S. venezuela* methymycin/pikromycin cluster was incorporated in the expression vector to drive the expression of foreign genes (the *cal*H of

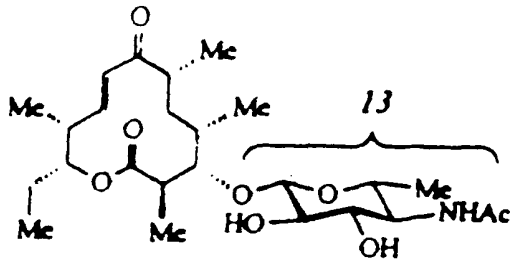
calicheamicin) in *S. venezuela.* In wild type *S. venezuela* methymycin/pikromycin pathway is known to produce methymycin, neomethymycin, pikromycin, and narbomycin. See figure 6. Deletion of the *des*1 gene in the mutant strain led to the accumulation of the CalH substrate, TDP-4-keto-6-deoxyglucose (compound 30, figure 6). The constructed expression vector with the *S. venezuela* promoter expressed the *cal*H gene to make the CalH protein. CalH acted on the substrate, 30, to produce compound 39 (figure 6). Compound 39 in turn, with the action of *S. venezuela's* DesVII (a glycosyltransferase) produced two methymycin/pikromycin-calicheamicin hybrid compounds. See Figure 6, compounds 40 and 41. These hybrid compounds carry the 4-aminohexose ligand of calicheamicin. This work provides indisputable support for the *cal*H gene assignment as encoding the TDP-6-deoxy -D-glycero-L-threo-4-hexulose 4-aminotransferase of the calicheamicin pathway. The CalH acted on the TDP-4-keto-deoxyglucose substrate (compound 30) to produce compound 39. (Figure 5).

Moreover, CalH is able to directly mediate the synthesis of the product TDP-4,6-dideoxy-alpha-D-glucose as demonstrated by HPLC isolation of the product and confirmation by high-resolution mass spectrometry. In addition this compound was found to co-elute with chemically synthesized TDP-4-amino-4,6-dideoxy-alpha-D-glucose.

In addition, these results reinforce the indiscriminate nature of the corresponding glycosyltransferase (DesVII) as they reveal that the glycosyltransferase (DesVII) of the *S. venezuela* pathway can recognize alternative sugar substrates whose structures are considerably different from the original amino sugar substrate, TDP-D-desosamine. The results also clearly demonstrate the ability to engineer secondary metabolite glycosylation through a rational selection of gene combinations. The successful expression of the CalH

protein in *S. venezuela* by the newly constructed expression vector highlights the potential of using this system to express other foreign genes in this strain.

Thus, one aspect of the present invention further relates to the construction of a composite gene cluster having the ability to make and attach non-natural sugars. The invention further provides an expression vector having a calicheamicin gene operably linked to regulatory sequences to control expression of the calicheamicin protein, and preferably the regulatory sequence is a *Streptomyces* promoter. The present invention also relates to two newly synthesized sugars, compound (11) and compound (12)(figure 7). Compound 11 has the formula:



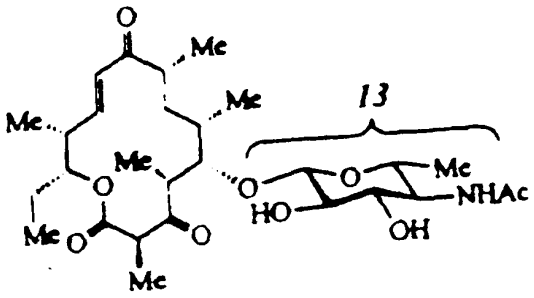
11

The spectral data of compound 11 was as follows:

^{1}H NMR (500 MHz $CDCl_3$, J in hertz) δ 6.75 (III, dd, J = 16.0, 5.5, 9-H) 6.44 (1H, dd, J = 16.0, 1.2, 8-H), 5.34 (1H, d, j = 8.0, N-H), 4.96 (1H, m, 11-H), 4.27 (1H, d, J=7.5, 1-H), 3.66 (1H, dd, J = 9.5, 8.0, 4'-H), 3.60 (1H, d, J = 10.5, 3-H), 3.50 (1H, 1, J - 9.5, 3'H), 3.^{d} (1H, m, 5'-H), 3.4 (1H, m, 2'-H), 2.84 (1H, dq, J = 10.5, 7.5, 2-H), 2.64 (1H, m, 10-H), 2.53 (1H, m, 6-H), 2.06 (3H, s, Me-C=0), 1.7 (1H, m, 12-H), 1.66 (1H, m, 5-H), 1.56 (1H, m. 12-H), 1.4 (1H, M, 5-H), 1.36 (3H, d., J=7.5, 2-Me), 1.25 (311. d, J = 6.5, 5'-

Me), 1.24 (1H, m. 4-H), 1.21 (3H, d, J=7.5, 6 Me), 1.10 (3H, d, J=6.5, 10-Me), 0.99 (3H, d, J=6.0, 4-Me), 0.91 (3H, t, J =7.2, 12-Me); ^{13}C NMR (125 MHz, $CDCl_3$) δ 205.3 (C-7), 175.1 (C-1), 171.9 (Me-C-O), 147.1 (C-9), 126.1 (C-8), 103.0 (C-1'), 85.8 (C-3), 75.8 (C-5'), 75.8 (C-3'), 74.1 (C-11) 70.8 (C-2'), 57.6 (C-4'), 45.3 (C-6), 44.0 (C-2), 38.1 (C-10), 34.2 (C-5), 33.6 (C-4), 25.4 (C-12), 23.7 (Me-C-O), 18.1 (C-6'), 17.9 (6 Me), 17.6 (4-Me), 16.4 (2-Me), 10.5 (12-Me), 9.8 (10-Me). High-resolution FAB-MS calculated for $C_{25}H_{42}$-NO_8 ($M + H^+$) 484.2910, found 484.2303.

Compound 12 has the formula:



12

The spectral data of compound 12 was as follows:

^1H NMR (500 MHz, $CDCl_3$. J in hertz) δ 6.69 (1H, dd, J = 16.0, 6.0, 11-H), 6.09 (1H, dd, J = 16.0, 1.5, 10-H), 5.35 (1H, d, J = 8.5, N-H), 4.96 (1H, m, 13-H), 4.36 (1H, d, J = 7.5, 1'H), 4.19 (1H, m. 5-H), 3.83 (1H-q, J=6.5, 2-H), 3.68 (1H, dt, J=10.0, 8.5, 4'H), 3.52 (1H, t, J = 8.5, 3-'H), 3.50 (1H, m, 5-H), 3.42 (1H, t, J = 7.5, 2'-H), 2.92 (1H, dq, J = 7.0, 5.0, 4-H), 2.81 (1H, m, 8-H), 2.73 (1H, t, J=7.5, 2'-H), 2.06 (3H, a, Me-C-O), 1.8 (1H, m, 6-H), 1.6 (1H, m, 14-H), 1.55 (1H, m. 7-H), 1.37 (3H, d, J = 6.5, 2-Me), 1.32 (3H, d,

J = 7.0, 4-Me), 1.3 (1H, m, H-14), 1.27 (3H, d, J = 6.5, 5'-Me), 1.25 (1H, m, 7-H), 1.12 (3H, d, J =6.0, 8-Me), 1.11 (3H, d, J = 6.5, 12-Me), 1.07 (3H, d, J = 6.0, 6-Me), 0.91 (3H, 1, J -7.2, 1 + Me); high resolution FAB MS calculated for C_{28} H_{46} NO_2 ($M+H^+$) 540.3172.found 540.3203.

One aspect of the invention relates to an isolated DNA strand containing the *cal*G gene and having the DNA sequence SEQ ID. NO.: 5. Another aspect of the invention is the protein, CalG, having amino acid sequence SEQ ID. No.: 6. According to BLAST analysis, *cal*G encodes a 4,6-dehydratase. Dehydratases had been characterized from *E. coli, Salmonella* and *Streptomyces,* (Thompson, M. et al., *J. Gen. Microbiol.*, 138, 779-786 (1992); Vara, J.A., et al., *J. Biol. Chem., 263,* 14992-14995 (1988)), and analogous NDP-D-glucose 4,6-dehydratases had been characterized from a variety of organisms. Liu, H.-w., et al., *Ann. Rev. Microbiol., 48,* 223-256 (1994); Hallis, T.M., et al., *Acc. Chem. Res.,* in press (1999). Based upon these prior studies, it was known that the overall transformation catalyzed by 4,6-dehydratases is an intramolecular oxidation-reduction where an enzyme-bound NAD^+ receives the 4-H as a hydride in the oxidative half-reaction and passes the reducing equivalents to C-6 of the dehydration product in the reductive half-reaction. Thus, it appears that Cal G is necessary for the formation of the aryltetrasaccharide 4,6-dideoxy-4-hydroxylamino-D-glucose moiety. CalG appears to be a TDP-D-glucose 4,6-dehydratase which catalyzes the conversion of intermediate 13 into intermediate 30. (See figure 5). Another aspect of the invention is an expression vector containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably, *E. coli*, containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule.

Moreover, CalG is able to directly mediate the synthesis of the product TDP-4-keto-6-deoxy-alpha-D-glucose as demonstrated by an assay where in the product is known to absorb at 320 nm under basic conditions. In addition this compound was found to co-elute with chemically synthesized TDP-4-keto-6-dideoxy-alpha-D-glucose. CalG has been demonstrated to utilize UDP-glucose as a substrate.

There is also disclosed an isolated DNA strand containing the *cal*S gene. Based on sequence homology with other P450-oxidases, CalS appears to be a P450-oxidase homolog which performs the oxidation of intermediate 39 to intermediate 42 (figure 5). The oxidation may occur at the nucleotide sugar level or hydroxylamine formation after the sugar has been transferred to the aglycone. There is also provided an expression vector containing the *cal*S gene or a fragment of *cal*S encoding for a bioactive molecule. There is also provided a transformed host cell, preferably bacteria, more preferably *E. coli*, containing *cal*G or a fragment of *cal*G encoding for a bioactive molecule.

There is also disclosed an isolated DNA strand containing the *calQ* gene. Based on sequence homology, CalQ appears to be a UDP-D-glucose-6 dehydrogenase homolog. The CalQ assay is based upon the requirement of this enzyme for two equivalents of NAD+ for activity. Thus, an assay based upon the increase in absorbance (as a result of the conversion of NAD+ to NADH upon the conversion of UDP-alpha-D-glucose to UDP-alpha-D-glucuronic acid). The product was also shown to co-elute with commercially available UDP-glucuronic acid and separately confirmed by high resolution mass spectrometry. This enzyme was also shown to utilize TDP-glucose.

There is also provided an expression vector containing the *calQ* gene or a fragment of *calQ* encoding for a bioactive molecule. There is also provided a transformed host cell,

preferably bacteria, more preferably *E. coli*, containing *calQ* or a fragment of *calQ* encoding for a bioactive molecule.

The present invention allows genetic manipulation of the biosynthetic gene cluster to produce calicheamicin analogs. The present invention provides for producing calicheamicin analogs by constructing deletions or substitutions of the genes involved in biosynthesis of the aryltetrasaccharide. The invention further provides for *in vitro* glycosylation by altering the glycosylation pattern of calicheamicin (via a glycosyltransferase) to produce additional analogs. The invention also provides for alteration of the calicheamicin aglycone by genetic manipulation of the genes encoding the biosynthesis of the warhead. Genetic manipulation, such as producing deletions or substitutions are performed using methods known in the art.

The invention provides for a method of purifying calicheamicin through affinity chromatography. Because of its homology with calicheamicin, CalC functions as a calicheamicin-sequestering/binding protein. Affinity chromatography is performed using methods known in the art.

The invention relates to the expression of the genes located in the biosynthetic gene cluster by using methods known in the art to insert the genes into a suitable expression vector and operably linking the gene to regulatory sequences to control expression of the gene to produce the protein encoded by the inserted gene. The present invention also provides for expression of biologically active proteins by inserting fragments of genes selected from the biosynthetic gene cluster, which encode for biologically active proteins, into a suitable expression vector, using methods known in the art. The genes would be operably linked to regulatory sequences to control their expression.

The term "hybridization" as used herein is generally used to mean hybridization of nucleic acids at appropriate conditions of stringency as would be readily evident to those skilled in the art depending upon the nature of the probe sequence and target sequences. Conditions of hybridization and washing are well known in the art, and the adjustment of conditions depending upon the desired stringency by varying incubation time, temperature and/or ionic strength of the solution are readily accomplished. See, for example, Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989. The choice of conditions is dictated by the length of the sequences being hybridized, in particular, the length of the probe sequence, the relative G-C content of the nucleic acids and the amount of mismatches to be permitted. Low stringency conditions are preferred when partial hybridization between strands that have lesser degrees of complementarity is desired. When perfect or near perfect complementarity is desired, high stringency conditions are preferred. For typical high stringency conditions, the hybridization solution contains 6x S.S.C., 0.01 M EDTA, 1x Denhardt's solution and 0.5% SDS. Hybridization is carried out at about 68°C for about 3 to 4 hours for fragments of cloned DNA and for about 12 to about 16 hours for total eukaryotic DNA. For lower stringencies the temperature of hybridization is reduced to about 12°C below the melting temperature (TM) of the duplex. The TM is known to be a function of the G-C content and duplex length as well as the ionic strength of the solution.

As used herein, the term "substantial sequence identity" or "substantial homology" is used to indicate that a nucleotide sequence or an amino acid sequence exhibits substantial structural or functional equivalence with another nucleotide or amino acid sequence. Any structural or functional differences between sequences having substantial

sequence identity or substantial homology will be *de minimis*; that is, they will not substantially affect the ability of the sequence to function as indicated in the desired application. Differences may be due to inherent variations in codon usage among different species, for example. Structural differences are considered de minimis if there is a significant amount of sequence overlap or similarity between two or more different sequences or if the different sequences exhibit similar physical characteristics even if the sequences differ in length or structure. Such characteristics include for example, ability to hybridize under defined conditions, or in the case of proteins, immunological crossreactivity, similar enzymatic activity, etc.

Additionally, two nucleotide sequences are "substantially complementary" if the sequences have at least about 40 percent, more preferably, at least about 60 percent and most preferably about 90 percent sequence similarity between them. Two amino acid sequences are "substantially homologous" if they have at least 40%, preferably 70% similarity between the active portions of the polypeptides.

As used herein, the phrase "hybridizes to a corresponding portion" of a DNA or RNA molecule means that the molecule that hybridizes, e.g., oligonucleotide, polynucleotide, or any nucleotide sequence (in sense or antisense orientation) recognizes and hybridizes to a sequence in another nucleic acid molecule that is of approximately the same size and has enough sequence similarity thereto to effect hybridization under appropriate conditions. It is to be understood that the size of the "corresponding portion" will allow for some mismatches in hybridization such that the "corresponding portion" may be smaller or larger than the molecule which hybridizes to it, for example 20-30% larger or smaller, preferably no more than about 12-15 % larger or smaller.

The term "functional derivative" of a nucleotide sequence (or poly- or oligonucleotide) is used herein to mean a fragment, variant, homolog, or analog of the nucleotide sequence of interest or of the nucleotide sequence encoding the peptide of interest. A functional derivative may include alternative codons for amino acids, or may code for different amino acids which do not substantially change the function of interest of the peptide encoded by the nucleotide. A functional derivative may retain at least a portion of the function of the nucleotide sequence of interest or of the nucleotide sequence encoding the peptide of interest, which function permits its utility in accordance with the invention. Such function may include the ability to hybridize with at least one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93 , or 94; the ability to hybridize with a substantially homologous DNA from another organism which DNA encodes at least one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95 or a functional derivative thereof, or with an mRNA transcript thereof, or the ability to encode a protein that is a functional derivative of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95, or the like.

A "fragment" of the gene or nucleotide sequence refers to any subset of the molecule, e.g., a shorter polynucleotide or oligonucleotide. A "variant" refers to a molecule substantially similar to either the entire gene or a fragment thereof, such as a nucleotide substitution variant having one or more substituted nucleotides, but which

maintains the ability to hybridize with the particular gene or to encode mRNA transcript which hybridizes with the native DNA. A "homolog" refers to a fragment or variant sequence from a different genus or species. An "analog" refers to a non-natural molecule substantially similar to or functioning in relation to either the entire molecule, a variant or a fragment thereof.

"Functional derivatives" of the proteins as described herein are fragments, variants, analogs, or chemical derivatives of at least one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95, and which retain at least a portion of the activity of at least one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95 or retain immunological cross reactivity with an antibody specific for at least one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 and 95. As used herein, a fragment of the protein refers to any subset of the molecule. Variant peptides may be made by direct chemical synthesis, for example, using methods well known in the art. An analog of a protein refers to a non-natural protein substantially similar to either the entire protein or a fragment thereof. As used herein, a chemical derivative of a protein may contain additional chemical moieties not normally a part of the peptide or peptide fragment. Modifications may be introduced into the a peptide or fragment thereof by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues.

A protein or peptide according to the invention may be produced by culturing a cell transformed with a nucleotide sequence of this invention (in the sense orientation), allowing the cell to synthesize the protein and then isolating the protein, either as a free protein or as a fusion protein, depending on the cloning protocol used, from either the culture medium or from cell extracts. Alternatively, the protein can be produced in a cell-free system. Ranu, et al., Meth. Enzymol., 60:459-484, (1979).

As can be appreciated from the disclosure above, the present invention has a wide variety of applications. Accordingly, the following examples are offered by way of illustration, not by way of limitation.

EXAMPLES

Example 1

To rapidly elucidate the nucleotide sequence, thermocycle sequencing was accomplished from pUC- or pBluescript-based subclones (using M13 primers and primer walking) as well as directly from isolated cosmids (via primer walking). Nucleotide sequence data was acquired using two Applied Biosystems automated 310 genetic analyzers and sequences were subsequently assembled using the Applied Biosystems AutoAssembler™ DNA sequence assembly software. Dear, S., et al., *Nucl Acids Res., 14,* 3907-3911 (1991); Huang, X., *Genomics, 14,* 18-25 (1992). Orf assignments were accomplished using a combination of the computational programs MacVector™ 6.0 and Brujene. MacVector is a commercially available software package which provides the ability to construct a *Micromonospora* codon bias table (from known *Micromonospora* sequences) and subsequently use this codon bias table to search for optimal orfs. Fickett,

J.W., *Nucleic Acids Research, 10,* 5303-5318 (1982). Alternatively, the shareware program Brujene was specifically designed for streptomycetes and assigns priority to orfs that illustrate a consistency high G/C% in the wobble position.

Example 2: Isolating and Characterizing *calC*

To isolate the gene(s) responsible for calicheamicin resistance in *Micromonospora*, clones conferring calicheamicin resistance were selected by growth of a *Micromonospora* genomic bifunctional cosmid library on LB plates containing ampicillin (50 μg ml^{-1}) and calicheamicin (0.25 μg ml^{-1}). In this selection, six clones (3a, 4a, 4b, 10a, 13a and 16a) displayed resistance to calicheamicin. Restriction mapping of these clones localized the desired phenotype to a ~2kb *PstI-SacI* fragment of DNA. (Figure 2). Maximum tolerated concentrations of calicheamicin on the LB plates was ascertained. The results are as follows:

Cosmid or Plasmid	**Maximum tolerated concentration of calicheamicin**
cosmids 3a, 4a, 10a, 13a, and 16a	0.5 μg ml^{-1}
pJT1214 and pJT1232	5.0 μg ml^{-1}
pRE7	20.0 μg ml^{-1}
induced pRE7	50.0 μg ml^{-1}
pJT1224 , pAP6, Pre1, and control plasmids pUC18, pBluescript, and pMAL-	<0.01 μg ml^{-1}

C2	

Nucleotide sequence analysis of the *PstI-SacI* fragment suggested that it contained two possible orfs. The proximal 1 kb of this fragment carried the single orf *calD* while the distal 1 kb presented orf *calC*. Computer translation of *calC* and subsequent BLAST analysis revealed no homology with known proteins, while the translation of *calD* to its respective protein, CalD, revealed the presence of three amino acid motifs typically conserved in S-adenosylmethionein-utilizing O-methyltransferases. Therefore, it was hypothesized that *calD* was not responsible for calicheamicin resistance. To rule out *calD* as being responsible for calicheamicin resistance, a subclone was engineered (*pJT1224*) to contain an intact *calD*, but the truncated *calC* gene. This subclone was not able to confer resistance to calicheamicin. Next, a subclone containing the *calC* region was constructed (*pJT1232*). This clone conferred calicheamicin resistance, as indicated in the above chart.

To ascertain the amino acid sequence of CalC and learn its properties, *cal*C was cloned into a pMAL-C2 vector. (pMAL-C2 by itself could not confer calicheamicin resistance. See above chart.) The resulting plasmid, pRE7, which contained *calC*, conferred resistance to calicheamicin. See above chart. Plasmid pRE7 was then induced with isopropyl Beta-D-thiogalactoside ("IPTG") to overexpress CalC. Induced pRE7 conferred resistance to calicheamicin and produced a maltose-binding protein CalC fusion protein (mbp-CalC). This resulting overexpression of CalC increased calicheamicin resistance 10^2-fold *in vivo*. See above chart.

Example 3: Expression of protein CalC

The protein mbp-CalC was overexpressed and purified for further analysis. The mbp-CalC was purified from pRE7/*E. coli* to homogeneity as judged by SDS-PAGE. An overnight LB culture (containing 50 mg ml^{-1} ampicillin and 50 ng ml^{-1} calicheamicin from a fresh pRE7/*E. coli* colony was grown at 37 °C, 250 rpm to an A_{600}=0.5, induced with 0.5 mM IPTG and growth continued overnight. Cells were harvested (4,000 x g, 4 °C, 20 minutes), resuspended in buffer A (50mM Tris-Cl, pH 7.5, 200 mM NaCl, 1mM EDTA) and disrupted by sonication. The cell debris was removed by centrifugation (5,000xg, 4°C, 20 minutes). The supernatant was applied to an amylose affinity column (1.5 x 7.0 cm, 1 mL min^{-1}). The desired mbp-CalC protein was eluted with buffer A containing 10 mM maltose. The eluate was concentrated and chromatographed on an S-300 column (50mM Tris-Cl, pH 7.5, 200 mM NaCl). Active fractions were used immediately or frozen at -80°C for storage.

Example 4: Verification of CalC's calicheamicin resistance

Given that calicheamicin leads to double strand DNA cleavage and CalC provides calicheamicin-resistance *in vivo*, it was expected that the addition of CalC to an *in vitro* calicheamicin-induced DNA cleavage assay would inhibit DNA cleavage. To test this theory, preliminary assays were performed with supercoiled pBlusecript plasmid DNA ("pBS") as the template, and dithiothreitol ("DTT") as the reductive initiator. In a typical assay, purified mbp-CalC (15.0 nM) and 30.0 nM calicheamicin were preincubated for 15 min. in a total volume of 25 μL 40 mM Tris-Cl, pH 7.5, at 37 °C. Then 2.5 μL 10mM DTT stock solution was added to the assay solution, and the assay was incubated an

additional 1 hour at 37°C. DNA fragmentation was assessed by electrophoresis on a 1% agarose gel stained with ethidium bromide. Using this assay, it was found that mbp-CalC could completely inhibit calicheamicin-induced DNA cleavage at concentrations nearing 10^3-fold excess of calicheamicin. Preincubation of mbp-CalC and DTT, protein removal via forced dialysis, and the subsequent use of the DTT solution as reductant did not noticeably affect the amount of DNA cleavage.

As indicated in Figure 4(b), no DNA cleavage was observed in the absence of DTT or calicheamicin (lanes a and b), while efficient cleavage was demonstrated in the presence of DTT and calicheamicin (lane c). As expected, the addition of mbp-CalC completely inhibited calicheamicin-induced DNA cleavage (lane f) while the addition of mbp alone (lane d) as a control, failed to inhibit calicheamicin-induced DNA cleavage. Furthermore, preincubation of mbp-CalC with DTT (not shown), or *apo*-mbp-CalC (lacking the Fe cofactor)(lane e), also failed to inhibit calicheamicin-induced DNA cleavage. However, the addition of Fe^{+2} or Fe^{+3} to the *apo*-mbp-CalC assay could reconstitute CalC activity (lane g). Reconstitution of *apo*-mbp-CalC was accomplished by preincubation with 1 mM $FeSO_4$ (Fe^{+2}) or $FeCl_3$ (Fe^{+3}) prior to the activity assay as previously described.

Example 5: Production of methymycin/pikromycin-calicheamicin hybrid compounds

The 1.2 kb *cal*H gene was amplified by polymerase chain reaction (PCR) from pJST1192$_{kpn7}$, which is a subclone containing a 7.0 kb *Kpn*I fragment of cosmid 13a. The amplified gene was cloned into the *Eco*RI/*Xba*I site of the expression vector pDHS617. This expression vector contains an apramycin resistance marker. The plasmid pDHS617 was derived from pOJ1446 (Bierman, M. et al., *Gene* 1992, 116, 43-49). A promoter

sequence from the *S. venezuela* methymycin/pikromycin cluster was incorporated in the plasmid to drive the expression of foreign genes in *S. venezuela.* The resulting plasmid, pLZ-C242 (containing the *cal*H gene insert and the promoter sequence) was introduced by conjugal transfer using *E.coli* S 17-1 into a previously constructed *S. venezuela* mutant, desI. (Borisova, S. et al., *Org. Lett.* 1999. 1. 133-136). In the DesI mutant, the *des*I was replaced by the neomycin resistance gene, which confers resistance to kanamycin The PLS-C242-containing *S. venezuela*-DesI colonies were identified on the basis of their resistance to apramycin antibiotic. One of these positive colonies, DesI/calH-1 was grown in 100 ml of seed medium at 29°C for 48 hours and then inoculated and grown in five Liters of vegetative medium. Cane, D.E., et al., *J. Am. Chem. Soc.*, 1993, 115, 522-526. The culture was centrifuged to remove cellular debris and mycella. The supernatant was adjusted to pH 9.5 with concentrated KOH, followed by chloroform extraction. The crude products (700 mg) were subjected to flash chromatography on silica gel using a gradient of 1-20% methanol in chloroform. A major product, 10-deoxymethynolide (ca. 400 mg), and a mixture of two minor macrolide compounds were obtained. The two macrolides were further purified by HPLC on a C_{18} column using an isocratic mobile phase of acetonitrile/H_2O (1:1). They were later identified as compound (11) and compound (12)(figure 7) by spectral anaylses.

Example 6: Molecular Break Light Assay

The invention further provides for a method of assaying the calicheamicin-induced DNA cleavage and its CalC mediated inhibition using the molecular break light assay. Two molecular break lights for the experiments are shown in Fig. 13. Break light A was

comprised of a 10-base pair stem which contained the known calicheamicin recognition sequence 5'-TCCT-3', while break light B carried the *BamH*I endonuclease recognition sequence 5'-GGATCC-3'. The length of break light B also considered the requirement of a 3 base pair overhang required for *BamH*I recognition and the stem of break light A was adjusted to a comparable length and melting temperature. The loop of both probes consisted of a T_4 loop to ensure non-hybridizing interactions. The 5'-fluorophore of both probes was fluorescein (FAM, $absorbance_{max}$ = 485 nm, $emission_{max}$ = 517 nm) while the corresponding 3'-quencher was 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL). Previous studies have shown DABCYL to serve as a universal quencher in molecular beacons and there is significant spectral overlap (1.02×10^{-15} M^{-1} cm^3) between the emission spectrum of FAM and the absorption spectrum of DABCYL. In a typical molecular beacon, the quenching efficiency of this pair via FRET has been shown to be essentially complete (99.9%), providing a significant enhancement of the signal to noise ratio as compared to typical complementary oligonucleotide pair FRET-based assays.

Enzymatic Cleavage as Proof of Principle. The first test was to demonstrate the specificity of the designed molecular break lights via enzymatic cleavage. Specifically, only break light B should cleave in the presence of the restriction endonuclease *BamH*I while both A and B should be digested by the non-specific nuclease DNaseI. As anticipated, Fig. 14a reveals a time dependent and [*BamH*I]-dependent increase of fluorescence only with B while A shows no change at 37 °C. Fig. 14b illustrates an increase of fluorescence over time with either break light A or B when digested with DNaseI which is also [DNaseI]-dependent. In comparison, control samples containing break lights alone or break lights in the presence of BSA gave no change in fluorescence

over > 2 hr at 37 °C. Given the lack of fluorescence in the absence of enzyme, the designed break lights show no appreciable melting at the designated assay temperature. Furthermore, these experiments clearly demonstrate the specificity of cleavage by *BamH*I for B and, for the first time, illustrate the principle application of molecular break lights to assess DNA cleavage.

Interestingly, the fluorescence maximum intensity obtained upon complete *BamH*I cleavage was only 75% that observed in the presence of DNaseI at the same concentration of molecular break light. Furthermore, after the *BamH*I reaction was complete, the addition of *BamH*I showed no change while the addition of DNaseI resulted in additional cleavage to give the expected 100% fluorescence maximum. This observation suggests the poly-guanidine tail left attached to FAM upon *BamH*I digestion quenches the fluorescent signal by ~25%. Consistent with this finding, PAGE analysis of the reaction products confirmed the presence of a 3-base overhang after excess treatment with *BamH*I which is completely degraded upon DNaseI digestion. As a result, the fluorescence maxium observed with excess *BamH*I was designated 100% cleavage for the *BamH*I kinetic studies described below.

Enediyne-Catalyzed Cleavage. Previous assays for enediyne cleavage of DNA relied upon discontinuous assays using radioactive DNA probes, electrophoresis and subsequent phosphoimager analysis. In contrast, by using break lights one can directly follow the extent of DNA cleavage by a specific enediyne in real time with high sensitivity. To demonstrate, Fig. 15a,b and Fig. 16a,c,d illustrate cleavage of break light A with varying concentrations of either (**1**) naturally-occurring enediynes including esperamicin, (**2**), non-enediyne small molecule agents (such as bleomycin (**3**) methidiumpropyl-Fe-EDTA, (**4**),

and Fe-EDTA, (**5**)) as well as the restriction endonuclease *Bam*HI) in the presence of excess reductive activator DTT. Under the conditions described, this assay allows the detection of **1** in the pM range. This sensitivity compares to that of the biochemical induction assay (BIA), the method of choice in detecting DNA-damaging agents. Furthermore, the sensitivity can be significantly enhanced by simply increasing the concentration of the molecular break light in the assay as demonstrated with the iron-dependent agents. The observed maximum fluorescence obtained upon cleavage of 3.2 nM break light A with either **1** or **2** was identical to that observed with DNaseI, consistent with complete degradation of the oligonucleotide. As controls, incubation of molecular break light A with either DTT or enediyne alone revealed no change in fluorescence. Furthermore, although there is some debate regarding the "specificity" of **1**, molecular break light B was cleaved by **1** at an identical rate. This supports the view that the specificity of **1** is more dependent upon context and perhaps less so on DNA sequence. It should also be noted that **1** leads to predominately double-stranded cleavage while **2** provides single-stranded nicks and the current molecular break light assay can not distinguish these two phenomena.

Interestingly, two distinct rates were observed in the enediyne molecular break light assay. The first (0-50 seconds) is a lag time most likely attributed to the enediyne activation while the second (50-200 seconds) is indicative to the initial velocity of DNA cleavage. To confirm this, assays were also established in which DTT and enediyne were first preincubated for 1-5 min followed by initiation via the addition of the substrate oligonucleotide. In these preincubation experiments, the previously observed "lag time" attributed to activation was no longer evident while the initial velocity of DNA cleavage

was identical to that determined in the standard assay. Preincubation for longer periods (> 30 min) revealed the same phenomenon, suggesting "activated" enediynes are perhaps more stable in an aqueous aerobic environment than previously estimated.

CalC inhibits calicheamicin mediated DNA cleavage. As illustrated in figure 17, CalC directly inhibits of calicheamicin-mediated DNA cleavage in the break light assay. 3.6pM break light A is coincubated with 3.5nM calicheamicin with increasing amounts of CalC (0.0nm, 1.3nm, 2.6nm, 3.9nm, 5.2nm). Complete inhibition of calicheamicin is achieved with roughly 2-fold excess of CalC. CalC has no effect on esperamicin-induced cleavage of DNA (data not shown).

All publications, patents and patent applications referred to herein are incorporated in this application by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.